

**COUNTY OF SAN DIEGO
REDISTRICTING ADVISORY COMMITTEE
APRIL 20, 2011**

Agenda Item F. Communications Received

The attached correspondence has been submitted by the public to the Redistricting Advisory Committee and is not directly related to a specific item on the agenda. The attachment includes correspondence received from April 7, 2011 to April 14, 2011 at 5 p.m.

The only action for the Committee to take is to note and file the receipt of the correspondence.

Potter, Andrew

From: CSG, Redistricting 2011
Sent: Thursday, April 07, 2011 4:31 PM
To: Adam Day; Andrea Skorepa; Deanna Weeks; Dennis Ridz; Michel Anderson
Cc: Pettingill, William L; Potter, Andrew
Subject: Sampling error in the U.S. Census American Community Survey (ACS)

Redistricting Advisory Committee,

Based on an inquiry by Committee Member Andrea Skorepa, I am sending some additional information about sampling error in the U.S. Census American Community Survey (ACS). The U.S. Census Bureau has published a paper that discusses sample design, estimation methodology, and accuracy of the data for the multiyear ACS and has posted it on their website here:

http://www.census.gov/acs/www/Downloads/data_documentation/Accuracy/MultiyearACSAccuracyofData2009.pdf

The Census Bureau also provides some guidance for data users in the form of training presentations here:

http://www.census.gov/acs/www/guidance_for_data_users/training_presentations/. Click on the presentation titled "Things that May Affect Estimates from the American Community Survey" for more information on this topic.

Please feel free to contact me if I can be of additional assistance.

Sincerely,

Nicole Cretelle Temple, MBA

Redistricting Project Manager

County of San Diego

(619) 531-5276

redistricting2011@sdcounty.ca.gov



Please consider the environment before printing this e-mail

American Community Survey Multiyear Accuracy of the Data (3-year 2007-2009 and 5-year 2005-2009)

Introduction

This multiyear ACS Accuracy of the Data document pertains to both the 2007-2009 3-year ACS data products and the 2005-2009 5-year ACS data products. Differences will be noted where applicable.

The data contained in these data products are based on the American Community Survey (ACS) sample. For the 3-year data products interviews from January 1, 2007 through December 31, 2009 were used. For the 5-year data products, interviews from January 1, 2005 through December 31, 2009 were used. Data products were produced for 1-year estimates (2005, 2006, 2007, 2008 and 2009), in addition to this set of 3-year and 5-year estimates.

In general, ACS estimates are period estimates that describe the average characteristics of population and housing over a period of data collection. The 2007-2009 ACS estimates are averages over the period from January 1, 2007 to December 31, 2009, and the 2005-2009 ACS estimates from January 1, 2005 through December 31, 2009, respectively. Multiyear estimates cannot be used to say what is going on in any particular year in the period, only what the average value is over the full period.

The ACS sample is selected from all counties and county-equivalents in the United States. In 2006, the ACS began collection of data from sampled persons in group quarters (GQ) – for example, military barracks, college dormitories, nursing homes, and correctional facilities. Persons in group quarters are included with persons in housing units (HUs) in all 2007-2009 and 2005-2009 ACS estimates based on the total population.

The ACS, like any other statistical activity, is subject to error. The purpose of this documentation is to provide data users with a basic understanding of the ACS sample design, estimation methodology, and accuracy of the 2007-2009 and 2005-2009 ACS estimates. The ACS is sponsored by the U.S. Census Bureau, and is part of the 2010 Decennial Census Program.

Additional information on the design and methodology of the ACS, including data collection and processing, can be found at http://www.census.gov/acs/www/methodology/methodology_main/

The **Multiyear Accuracy of the Data from the Puerto Rico Community Survey** can be found at http://www.census.gov/acs/www/data_documentation/documentation_main/

Table of Contents

Data Collection	3
Sample Design	3
Weighting Methodology	5
Estimation Methodology for Multiyear Estimates.....	6
Confidentiality of the Data.....	7
Errors in the Data	8
Measures of Sampling Error	9
Calculation of Standard Errors.....	10
Sums and Differences of Direct Standard Errors.....	11
Ratios	12
Proportions/Percents	12
Percent Change	12
Products.....	12
Differences of Estimates for Overlapping Periods of Identical Length.....	13
Testing for Significant Differences.....	13
Examples of Standard Error Calculations.....	14
Control of Nonsampling Error	17
Issues with Approximating the Standard Error of Linear Combinations of Multiple Estimates..	21

Data Collection

The ACS employs three modes of data collection:

- Mailout/Mailback
- Computer Assisted Telephone Interview (CATI)
- Computer Assisted Personal Interview (CAPI)

The general timing of data collection is:

- | | |
|----------|--|
| Month 1: | Addresses determined to be mailable are sent a questionnaire via the U.S. Postal Service. |
| Month 2: | All mail non-responding addresses with an available phone number are sent to CATI. |
| Month 3: | A sample of mail non-responses without a phone number, CATI non-responses, and unmailable addresses are selected and sent to CAPI. |

Sample Design

Sampling rates are assigned independently at the census block level. A measure of size is calculated for each of the following governmental units:

- Counties
- Places (active, functioning governmental units)
- School Districts (elementary, secondary, and unified)
- American Indian Areas (including Tribal Subdivisions beginning in 2008)
- Minor Civil Divisions (MCDs) – Connecticut, Maine, Massachusetts, Michigan, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, and Wisconsin (these are the states where MCDs are active, functioning governmental units)
- Alaska Native Village Statistical Areas
- Hawaiian Homelands

Each block is then assigned the smallest measure of size (GUMOS) from the set of all governmental units it is a part of.

The measure of size for all geographic entities for all areas (except American Indian Areas) is an estimate of the number of occupied housing units in the area. This was calculated by multiplying the number of ACS addresses by an estimate of the occupancy rate from Census 2000 and the ACS at the block level. For American Indian Areas the measure of size is the estimated number of occupied housing units multiplied by the proportion of people reporting American Indian (alone or in combination) in Census 2000. A measure of size for each census tract (TRACTMOS) was also calculated in the same manner.

Table 2. 2005 Through 2009 Sampling Rates for the United States

Sampling Rate Category	2005 Sampling Rates	2006 Sampling Rates	2007 Sampling Rates	2008 Sampling Rates	2009 Sampling Rates
Blocks in smallest governmental units (GUMOS < 200)	10.0%	10.0%	10.0%	10.0%	10.0%
Blocks in smaller governmental units (200 ≤ GUMOS < 800)	6.9%	6.8%	6.7%	6.6%	6.5%
Blocks in small governmental units (800 ≤ GUMOS ≤ 1200)	3.6%	3.4%	3.3%	3.3%	3.3%
Blocks in large tracts (GUMOS > 1200, TRACTMOS ≥ 2000) where mailable addresses ≥ 75% and predicted levels of completed mail and CATI interviews prior to CAPI subsampling > 60%	1.6%	1.6%	1.5%	1.5%	1.5%
Other blocks in large tracts (GUMOS > 1200, TRACTMOS ≥ 2000)	1.7%	1.7%	1.6%	1.6%	1.6%
All other blocks (GUMOS > 1200, TRACTMOS < 2000) where mailable addresses ≥ 75% and predicted levels of completed mail and CATI interviews prior to CAPI subsampling > 60%	2.1%	2.1%	2.1%	2.0%	2.0%
All other blocks (GUMOS > 1200, TRACTMOS < 2000)	2.3%	2.3%	2.2%	2.2%	2.2%

Addresses determined to be unmailable do not go to the CATI phase of data collection and are subsampled for the CAPI phase of data collection at a rate of 2-in-3. Subsequent to CATI, all addresses for which no response has been obtained are subsampled. This subsample is sent to the CAPI data collection phase. Beginning with the CAPI sample for the January 2005 panel (March 2005 data collection), the CAPI subsampling rate was based on the expected rate of completed mail and CATI interviews at the tract level.

Table 3. 2005 Through 2009 CAPI Subsampling Rates for the United States

Address and Tract Characteristics	CAPI Subsampling Rates
Unmailable addresses and addresses in Remote Alaska	66.7%
Mailable addresses in tracts with predicted levels of completed mail and CATI interviews prior to CAPI subsampling between 0% and less than 36%	50.0%
Mailable addresses in tracts with predicted levels of completed mail and CATI interviews prior to CAPI subsampling greater than 35% and less than 51%	40.0%
Mailable addresses in other tracts	33.3%

For a more detailed description of the ACS sampling methodology, see the 2009 ACS Accuracy of the Data document

(http://www.census.gov/acs/www/Downloads/data_documentation/Accuracy/ACS_Accuracy_of_Data_2009.pdf). For more information relating to sampling in a specific year, please refer to the individual year's Accuracy of the Data document

http://www.census.gov/acs/www/data_documentation/documentation_main/.

Weighting Methodology

The multiyear estimates should be interpreted as estimates that describe a time period rather than a specific reference year. For example, a 3-year estimate for the poverty rate of a given area describes the total set of people who lived in that area over those three years much the same way as a 1-year estimate for the same characteristic describes the set of people who lived in that area over one year. The only fundamental difference between the estimates is the number of months of collected data which are considered in forming the estimate. For this reason, the estimation procedure used for the multiyear estimates is an extension of the 2009 1-year estimation procedure. In this document only the procedures that are unique to the multiyear estimates are discussed.

To weight the 3-year estimates, 36 months of collected data are pooled together and for the 5-year estimates, 60 months were pooled. The pooled data are then reweighted using the procedures developed for the 2009 1-year estimates with a few adjustments. These adjustments concern geography, month-specific weighting steps, and population and housing unit controls. In addition to these adjustments, there is one multiyear specific model-assisted weighting step.

Some of the weighting steps use the month of tabulation in forming the weighting cells within which the weighting adjustments are made. One such example is the non-interview adjustment. In these weighting steps, the month of tabulation is used independently of year. Thus, for the 3-year, sample cases from May 2007, May 2008, and May 2009 are combined into one weighting cell and for the 5-year, sample cases from May 2005, May 2006, May 2007, May 2008, and May 2009 are combined.

Since the multiyear estimates represent estimates for the period, the controls are not a single year's housing or population estimates from the Population Estimates Program, but rather are an average of these estimates over the period. For the housing unit controls, a simple average of the

1-year housing unit estimates over the period is calculated for each county. The version or vintage of estimates used is always the last year of the period since these are considered to be the most up-to-date and are created using a consistent methodology. For example, the housing unit control used for a given county in the 2005-2009 weighting is equal to the simple average of the 2005, 2006, 2007, 2008, and 2009 estimates that were produced using the 2009 methodology (the 2009 vintage). Likewise, the population controls by race, ethnicity, age, and sex are obtained by taking a simple average of the 1-year population estimates of the county by race, ethnicity, age, and sex. For example, the 2005-2009 control total used for Hispanic males age 20-24 in a given county would be obtained by averaging the 1-year population estimates for that demographic group for 2005, 2006, 2007, 2008, and 2009. The version or vintage of estimates used is always that of the last year of the period since these are considered to be the most up to date and are created using a consistent methodology.

One multiyear specific step is a model-assisted (generalized regression or GREG) weighting step. The objective of this additional step is to reduce the variances of base demographics at the place and MCD level in the 3-year estimates and at the tract level in the 5-year estimates. While reducing the variances, the estimates themselves are relatively unchanged. This process involves linking administrative record data with ACS data.

In addition, a finite population correction (FPC) factor is included in the creation of the replicate weights for both the 3-year and 5-year data at the tract level. It reduces the estimate of the variance and the margin of error by taking the sampling rate into account. A two-tiered approach was used. One FPC was calculated for mail and CATI respondents and another for CAPI respondents. The CAPI was given a separate FPC to take into account the fact that CAPI respondents are subsampled. The FPC is not included in the 1-year data because the sampling rates are relatively small and thus the FPC does not have an appreciable impact on the variance.

For more information on the replicate weights and replicate factors, see the Design and Methodology Report located at

http://www.census.gov/acs/www/methodology/methodology_main/.

Estimation Methodology for Multiyear Estimates

For the 1-year estimation, the tabulation geography for the data is based on the boundaries defined on January 1 of the tabulation year, which is consistent with the tabulation geography used to produce the population estimates. All sample addresses are updated with this geography prior to weighting. For the multiyear estimation, the tabulation geography for the data is referenced to the final year in the multiyear period. For example, the 2007-2009 period uses the 2009 reference geography. Thus, all data collected over the period of 2007-2009 in the blocks that are contained in the 2009 boundaries for a given place are tabulated as though they were a part of that place for the entire period.

Monetary values for the ACS 3-year estimates are inflation-adjusted to the final year of the period. For example, the 2007-2009 ACS 3-year estimates are tabulated using 2009-adjusted dollars. These adjustments use the national Consumer Price Index (CPI) since a regional-based

CPI is not available for the entire country. The ACS 5-year estimates are also inflation-adjusted in the same manner.

For a more detailed description of the ACS estimation methodology, see the 2009 Accuracy of the Data document (http://www.census.gov/acs/www/Downloads/data_documentation/Accuracy/ACS_Accuracy_of_Data_2009.pdf). For more information relating to estimation in a specific year, please refer to that individual year's Accuracy of the Data document (http://www.census.gov/acs/www/data_documentation/documentation_main/).

Confidentiality of the Data

The Census Bureau has modified or suppressed some data on this site to protect confidentiality. Title 13 United States Code, Section 9, prohibits the Census Bureau from publishing results in which an individual's data can be identified.

The Census Bureau's internal Disclosure Review Board sets the confidentiality rules for all data releases. A checklist approach is used to ensure that all potential risks to the confidentiality of the data are considered and addressed.

- Title 13, United States Code: Title 13 of the United States Code authorizes the Census Bureau to conduct censuses and surveys. Section 9 of the same Title requires that any information collected from the public under the authority of Title 13 be maintained as confidential. Section 214 of Title 13 and Sections 3559 and 3571 of Title 18 of the United States Code provide for the imposition of penalties of up to five years in prison and up to \$250,000 in fines for wrongful disclosure of confidential census information.
- Disclosure Limitation: Disclosure limitation is the process for protecting the confidentiality of data. A disclosure of data occurs when someone can use published statistical information to identify an individual that has provided information under a pledge of confidentiality. For data tabulations the Census Bureau uses disclosure limitation procedures to modify or remove the characteristics that put confidential information at risk for disclosure. Although it may appear that a table shows information about a specific individual, the Census Bureau has taken steps to disguise or suppress the original data while making sure the results are still useful. The techniques used by the Census Bureau to protect confidentiality in tabulations vary, depending on the type of data.
- Data Swapping: Data swapping is a method of disclosure limitation designed to protect confidentiality in tables of frequency data (the number or percent of the population with certain characteristics). Data swapping is done by editing the source data or exchanging records for a sample of cases when creating a table. A sample of households is selected and matched on a set of selected key variables with households in neighboring geographic areas that have similar characteristics (such as the same number of adults and same number of children). Because the swap often occurs within a neighboring area, there is no effect on the marginal totals for the area or for totals that include data from multiple areas. Because of data swapping, users should not assume that tables with cells

having a value of one or two reveal information about specific individuals. Data swapping procedures were first used in the 1990 Census, and were used again in Census 2000.

The data use the same disclosure limitation methodology as the original 1-year data. The confidentiality edit was previously applied to the raw data files when they were created to produce the 1-year estimates and these same data files with the original confidentiality edit were used to produce the 3-year and 5-year estimates.

Errors in the Data

- **Sampling Error** — The data in the ACS products are estimates of the actual figures that would have been obtained by interviewing the entire population using the same methodology. The estimates from the chosen sample also differ from other samples of housing units and persons within those housing units. Sampling error in data arises due to the use of probability sampling, which is necessary to ensure the integrity and representativeness of sample survey results. The implementation of statistical sampling procedures provides the basis for the statistical analysis of sample data.
- **Nonsampling Error** — In addition to sampling error, data users should realize that other types of errors may be introduced during any of the various complex operations used to collect and process survey data. For example, operations such as data entry from questionnaires and editing may introduce error into the estimates. Another source is through the use of controls in the weighting. The controls are designed to mitigate the effects of systematic undercoverage of certain groups who are difficult to enumerate and to reduce the variance. The controls are based on the population estimates extrapolated from the previous census. Errors can be brought into the data if the extrapolation methods do not properly reflect the population. However, the potential risk from using the controls in the weighting process is offset by far greater benefits to the ACS estimates. These benefits include reducing the effects of a larger coverage problem found in most surveys, including the ACS, and the reduction of standard errors of ACS estimates. These and other sources of error contribute to the nonsampling error component of the total error of survey estimates. Nonsampling errors may affect the data in two ways. Errors that are introduced randomly increase the variability of the data. Systematic errors which are consistent in one direction introduce bias into the results of a sample survey. The Census Bureau protects against the effect of systematic errors on survey estimates by conducting extensive research and evaluation programs on sampling techniques, questionnaire design, and data collection and processing procedures. In addition, an important goal of the ACS is to minimize the amount of nonsampling error introduced through nonresponse for sample housing units. One way of accomplishing this is by following up on mail nonrespondents during the CATI and CAPI phases.

Measures of Sampling Error

Sampling error is the difference between an estimate based on a sample and the corresponding value that would be obtained if the estimate were based on the entire population (as from a census). Note that sample-based estimates will vary depending on the particular sample selected from the population. Measures of the magnitude of sampling error reflect the variation in the estimates over all possible samples that could have been selected from the population using the same sampling methodology.

Estimates of the magnitude of sampling errors – in the form of margins of error – are provided with all published ACS estimates. The Census Bureau recommends that data users incorporate this information into their analyses, as sampling error in survey estimates could impact the conclusions drawn from the results.

Confidence Intervals and Margins of Error

Confidence Intervals – A sample estimate and its estimated standard error may be used to construct confidence intervals about the estimate. These intervals are ranges that will contain the average value of the estimated characteristic that results over all possible samples, with a known probability.

For example, if all possible samples that could result under the ACS sample design were independently selected and surveyed under the same conditions, and if the estimate and its estimated standard error were calculated for each of these samples, then:

1. Approximately 68 percent of the intervals from one estimated standard error below the estimate to one estimated standard error above the estimate would contain the average result from all possible samples;
2. Approximately 90 percent of the intervals from 1.645 times the estimated standard error below the estimate to 1.645 times the estimated standard error above the estimate would contain the average result from all possible samples.
3. Approximately 95 percent of the intervals from two estimated standard errors below the estimate to two estimated standard errors above the estimate would contain the average result from all possible samples.

The intervals are referred to as 68 percent, 90 percent, and 95 percent confidence intervals, respectively.

Margin of Error – Instead of providing the upper and lower confidence bounds in published ACS tables, the margin of error is provided instead. The margin of error is the difference between an estimate and its upper or lower confidence bound. Both the confidence bounds and the standard error can easily be computed from the margin of error. All ACS published margins of error are based on a 90 percent confidence level.

$$\text{Standard Error} = \text{Margin of Error} / 1.645$$

Lower Confidence Bound = Estimate - Margin of Error

Upper Confidence Bound = Estimate + Margin of Error

When constructing confidence bounds from the margin of error, the user should be aware of any “natural” limits on the bounds. For example, if a population estimate is near zero, the calculated value of the lower confidence bound may be negative. However, a negative number of people does not make sense, so the lower confidence bound should be reported as zero instead. However, for other estimates such as income, negative values do make sense. The context and meaning of the estimate must be kept in mind when creating these bounds. Another of these natural limits would be 100% for the upper bound of a percent estimate.

If the margin of error is displayed as ‘*****’ (five asterisks), the estimate has been controlled to be equal to a fixed value and so has no sampling error. When using any of the formulas in the following section, use a standard error of zero for these controlled estimates.

Limitations –The user should be careful when computing and interpreting confidence intervals.

- The estimated standard errors (and thus margins of errors) included in these data products do not include portions of the variability due to nonsampling error that may be present in the data. In particular, the standard errors do not reflect the effect of correlated errors introduced by interviewers, coders, or other field or processing personnel. Nor do they reflect the error from imputed values due to missing responses. Thus, the standard errors calculated represent a lower bound of the total error. As a result, confidence intervals formed using these estimated standard errors may not meet the stated levels of confidence (i.e., 68, 90, or 95 percent). Thus, some care must be exercised in the interpretation of the data in this data product based on the estimated standard errors.
- Zero or small estimates; very large estimates — The value of almost all ACS characteristics is greater than or equal to zero by definition. For zero or small estimates, use of the method given previously for calculating confidence intervals relies on large sample theory, and may result in negative values which for most characteristics are not admissible. In this case the lower limit of the confidence interval is set to zero by default. A similar caution holds for estimates of totals close to a control total or estimated proportions near one, where the upper limit of the confidence interval is set to its largest admissible value. In these situations the level of confidence of the adjusted range of values is less than the prescribed confidence level.

Calculation of Standard Errors

Direct estimates of the standard errors were calculated for all estimates reported in this product. The standard errors, in most cases, are calculated using a replicate-based methodology that takes into account the sample design and estimation procedures. Excluding the base weight, replicate weights were allowed to be negative in order to avoid underestimating the standard error. Exceptions include:

1. The estimate of the number or proportion of people, households, families, or housing units in a geographic area with a specific characteristic is zero. A special procedure is used to estimate the standard error.
2. There are either no sample observations available to compute an estimate or standard error of a median, an aggregate, a proportion, or some other ratio, or there are too few sample observations to compute a stable estimate of the standard error. The estimate is represented in the tables by “-” and the margin of error by “**” (two asterisks).
3. The estimate of a median falls in the lower open-ended interval or upper open-ended interval of a distribution. If the median occurs in the lowest interval, then a “-” follows the estimate, and if the median occurs in the upper interval, then a “+” follows the estimate. In both cases the margin of error is represented in the tables by “***” (three asterisks).

Sums and Differences of Direct Standard Errors

The standard errors estimated from these tables are for individual estimates. Additional calculations are required to estimate the standard errors for sums of or the differences between two or more sample estimates.

The standard error of the sum of two sample estimates is the square root of the sum of the two individual standard errors squared plus a covariance term. That is, for standard errors $SE(\hat{X}_1)$ and $SE(\hat{X}_2)$ of estimates \hat{X}_1 and \hat{X}_2 :

$$SE(\hat{X}_1 \pm \hat{X}_2) = \sqrt{(SE(\hat{X}_1))^2 + (SE(\hat{X}_2))^2 \pm covariance} \quad (1)$$

The covariance measures the interactions between two estimates. Currently the covariance terms are not available. Data users should use the approximation:

$$SE(\hat{X}_1 \pm \hat{X}_2) \approx \sqrt{(SE(\hat{X}_1))^2 + (SE(\hat{X}_2))^2} \quad (2)$$

However, this method will underestimate or overestimate the standard error if the two estimates interact in either a positive or negative way.

The approximation formula (2) can be expanded to more than two estimates by adding in the individual standard errors squared inside the radical. As the number of estimates involved in the sum or difference increases, the results of formula (2) become increasingly different from the standard error derived directly from the ACS microdata. Users are encouraged to work with the fewest number of estimates possible. If there are estimates involved in the sum that are controlled in the weighting then the approximate standard error can be increasingly different. Several examples are provided starting on page 21 to demonstrate issues associated with approximating the standard errors when summing large numbers of estimates together.

Ratios

The statistic of interest may be the ratio of two estimates. First is the case where the numerator *is not* a subset of the denominator. The standard error of this ratio between two sample estimates is approximated as:

$$SE\left(\frac{\hat{X}}{\hat{Y}}\right) = \frac{1}{\hat{Y}} \sqrt{[SE(\hat{X})]^2 + \frac{\hat{X}^2}{\hat{Y}^2} [SE(\hat{Y})]^2} \quad (3)$$

Proportions/Percents

For a proportion (or percent), a ratio where the numerator *is* a subset of the denominator, a slightly different estimator is used. If $\hat{P} = \hat{X}/\hat{Y}$, then the standard error of this proportion is approximated as:

$$SE(\hat{P}) = \frac{1}{\hat{Y}} \sqrt{[SE(\hat{X})]^2 - \frac{\hat{X}^2}{\hat{Y}^2} [SE(\hat{Y})]^2} \quad (4)$$

If $\hat{Q} = 100\% \times \hat{P}$ (P is the proportion and Q is its corresponding percent), then $SE(\hat{Q}) = 100\% \times SE(\hat{P})$. Note the difference between the formulas to approximate the standard error for proportions (4) and ratios (3) - the plus sign in the previous formula has been replaced with a minus sign. If the value under the radical is negative, use the ratio standard error formula above, instead.

Percent Change

Calculating the percent change from one time period to another. For example, computing the percent change of a 2008-2010 estimate to a 2007-2009 estimate. Normally, the current estimate is compared to the older estimate.

Let the current estimate = X and the earlier estimate = Y, then the formula for percent change is:

$$SE\left(\frac{\hat{X} - \hat{Y}}{\hat{Y}} \times 100\%\right) = 100\% \times SE\left(\frac{\hat{X}}{\hat{Y}} - 1\right) = 100\% \times SE\left(\frac{\hat{X}}{\hat{Y}}\right) \quad (5)$$

This reduces to a ratio. The ratio formula (3) above may be used to calculate the standard error. As a caveat, this formula does not take into account the correlation when calculating overlapping time periods.

Products

For a product of two estimates - for example if you want to estimate a proportion's numerator by multiplying the proportion by its denominator - the standard error can be approximated as:

$$SE(\hat{X} \times \hat{Y}) = \sqrt{\hat{X}^2 \times [SE(\hat{Y})]^2 + \hat{Y}^2 \times [SE(\hat{X})]^2} \quad (6)$$

Differences of Estimates for Overlapping Periods of Identical Length

For example, \hat{X} may represent an estimate of a characteristic for the period 2006-2008 and \hat{Y} the estimate of the same characteristic for 2007-2009. In this case, data for 2007 and 2008 are included in both estimates, and their contribution is largely subtracted out when differences are calculated. In this case, it is possible to approximate the sampling correlation between the two estimates to improve upon the previous expression, namely:

$$SE(\hat{X} - \hat{Y}) = \sqrt{1 - C} \sqrt{[SE(\hat{X})]^2 + [SE(\hat{Y})]^2} \quad (7)$$

where C is the fraction of overlapping years. For example, the periods 2006-2008 and 2007-2009 overlap by two out of three years, so $C = 2 / 3$ and $1 - C = 0.33$. If the periods do not overlap, such as 2005-2007 and 2008-2010, then no factor is needed. Due to the difficulty in interpreting overlapping time periods, the Census Bureau currently discourages users from making such comparisons.

Testing for Significant Differences

Significant differences – Users may conduct a statistical test to see if the difference between an ACS estimate and any other chosen estimates is statistically significant at a given confidence level. “Statistically significant” means that the difference is not likely due to random chance alone. With the two estimates (Est_1 and Est_2) and their respective standard errors (SE_1 and SE_2), calculate a Z statistic:

$$Z = \frac{Est_1 - Est_2}{\sqrt{(SE_1)^2 + (SE_2)^2}} \quad (8)$$

If $Z > 1.645$ or $Z < -1.645$, then the difference can be said to be statistically significant at the 90 percent confidence level.¹ Any estimate can be compared to an ACS estimate using this method, including other ACS estimates from the current year, the ACS estimate for the same characteristic and geographic area but from a previous year, Census 2000 100 percent counts and long form estimates, estimates from other Census Bureau surveys, and estimates from other sources. Not all estimates have sampling error — Census 2000 100 percent counts do not, for example, although Census 2000 long form estimates do — but they should be used if they exist to give the most accurate result of the test.

Users are also cautioned to *not* rely on looking at whether confidence intervals for two estimates overlap or not to determine statistical significance, because there are circumstances where that method will not give the correct test result. If two confidence intervals do not overlap, then the estimates will be significantly different (i.e. the significance test will always agree). However, if two confidence intervals do overlap, then the estimates may or may not be significantly different. The Z calculation above is recommended in all cases.

¹ The ACS Accuracy of the Data document in 2005 used a Z statistic of +/-1.65. Data users should use +/-1.65 for estimates published in 2005 or earlier.

Here is a simple example of why it is not recommended to use the overlapping confidence bounds rule of thumb as a substitute for a statistical test.

Let: $X_1 = 5.0$ with $SE_1 = 0.2$ and $X_2 = 6.0$ with $SE_2 = 0.5$.

The Upper Bound for $X_1 = 5.0 + 0.2 * 1.645 = 5.3$ while the Lower Bound for $X_2 = 6.0 - 0.5 * 1.645 = 5.2$. The confidence bounds overlap, so, the rule of thumb would indicate that the estimates are not significantly different at the 90% level.

However, if we apply the statistical significance test we obtain:

$$Z = \frac{5 - 6}{\sqrt{0.2^2 + 0.5^2}} = 1.857$$

$Z = 1.857 > 1.645$ which means that the difference is significant (at the 90% level).

All statistical testing in ACS data products is based on the 90 percent confidence level. Users should understand that all testing was done using *unrounded* estimates and standard errors, and it may not be possible to replicate test results using the rounded estimates and margins of error as published.

Examples of Standard Error Calculations

We will present some examples based on the real data to demonstrate the use of the formulas. All of the data used here are from 2005-2009 5-year data, but the process would be the same is 2007-2009 3-year data were used instead.

Example 1 - Calculating the Standard Error from the Confidence Interval

The estimated number of males, never married is 39,992,869 from summary table B12001 for the period 2005-2009 in the United States. The margin of error is 159,222.

Standard Error = Margin of Error / 1.645

Calculating the standard error using the margin of error, we have:

$SE(39,992,869) = 159,222 / 1.645 = 96,791$.

Example 2 - Calculating the Standard Error of a Sum

We are interested in the number of people who have never married for the period 2005-2009 in the United States. From example 1, we know the number of males, never married is 39,992,869. From summary table B12001 we have the number of females, never married is 34,028,805 with a margin of error of 120,787. So, the estimated number of people who have never been married is $39,992,869 + 34,028,805 = 74,021,674$. To calculate the standard error of this sum, we need the standard errors of the two estimates

in the sum. We have the standard error for the number of males never married from example 1 as 96,791. The standard error for the number of females never married is calculated using the margin of error:

$$SE(34,028,805) = 120,787 / 1.645 = 73,427.$$

So using the formula for the standard error of a sum or difference we have:

$$SE(74,021,674) = \sqrt{96,791^2 + 73,427^2} = 121,491.$$

Caution: This method, however, will underestimate (overestimate) the standard error if the two items in a sum are highly positively (negatively) correlated or if the two items in a difference are highly negatively (positively) correlated.

To calculate the lower and upper bounds of the 90 percent confidence interval around 74,021,674 using the standard error, simply multiply 121,491 by 1.645, then add and subtract the product from 74,021,674. Thus the 90 percent confidence interval for this estimate is $[74,021,674 - 1.645(121,491)]$ to $[74,021,674 + 1.645(121,491)]$ or 73,821,821 to 74,221,527.

Example 3 - Calculating the Standard Error of a Percent

We are interested in the percentage of females who have never married to the number of people who have never married during the period of 2005-2009. The number of females, never married is 34,028,805, and the number of people who have never married is 74,021,674. To calculate the standard error of this sum, we need the standard errors of the two estimates in the sum. We have the standard error for the number of females never married from example 2 as 73,427 and the standard error for the number of people never married calculated from example 2 as 121,491.

$$\text{The estimate is } (34,028,805 / 74,021,674) * 100\% = 45.97\%$$

So, using the formula for the standard error of a proportion or percent, we have:

$$SE(45.97\%) = 100\% * \frac{1}{74,021,674} \sqrt{73,427^2 - \frac{34,028,805^2}{74,021,674^2} \times 121,491^2} = 0.06$$

To calculate the lower and upper bounds of the 90 percent confidence interval around 45.97 using the standard error, simply multiply 0.06 by 1.645, then add and subtract the product from 45.97. Thus the 90 percent confidence interval for this estimate is $[45.97 - 1.645(0.06)]$ to $[45.97 + 1.645(0.06)]$, or 45.87% to 46.07%.

Example 4 - Calculating the Standard Error of a Ratio

Now, let us calculate the estimate of the ratio of the number of unmarried males to the number of unmarried females and its standard error. From the above examples, the estimate for the number of unmarried men is 39,992,869 with a standard error of 96,791, and the estimates for the number of unmarried women is 34,028,805 with a standard error of 73,427.

The estimate of the ratio is $39,992,869 / 34,028,805 = 1.175$.

The standard error of this ratio is

$$SE(1.175) = \frac{1}{34,028,805} \sqrt{73,427^2 + 1.175^2 \times 96,791^2} = 0.00398$$

The 90 percent margin of error for this estimate would be 0.00398 multiplied by 1.645, or about 0.007. The 90 percent lower and upper 90 percent confidence bounds would then be $[1.175 - 0.007]$ to $[1.175 + 0.007]$, or 1.168 and 1.182.

Example 5 - Calculating the Standard Error of a Product

We are interested in the number of 1-unit detached owner-occupied housing units in the U.S. The number of owner-occupied housing units is 75,320,422 with a margin of error of 342,226 from subject table S2504 for 2009, and the percent of 1-unit detached owner-occupied housing units is 81.6% (0.816) with a margin of error of 0.1% (0.001). So the number of 1-unit detached owner-occupied housing units is $75,320,422 \times 0.816 = 61,461,464$. Calculating the standard error for the estimates using the margin of error we have:

$$SE(75,320,422) = 342,226 / 1.645 = 208,040$$

and

$$SE(0.816) = 0.001 / 1.645 = 0.0006079$$

The standard error for number of 1-unit detached owner-occupied housing units is calculated using the formula for products as:

$$SE(61,461,464) = \sqrt{75,320,422^2 \times 0.0006079^2 + 0.816^2 \times 208,040^2} = 175,827$$

To calculate the lower and upper bounds of the 90 percent confidence interval around 61,461,464 using the standard error, simply multiply 175,827 by 1.645, then add and subtract the product from 61,461,464. Thus, the 90 percent confidence interval for this estimate is $[61,461,464 - 1.645(175,827)]$ to $[61,461,464 + 1.645(175,827)]$ or 61,172,229 to 61,750,699.

Example 6 - Calculating the Standard Error of the Difference of Overlapping Periods of Identical Length:

It should be noted that due to the difficulty in interpreting the “difference” in overlapping period estimates, the Census Bureau currently discourages users from making such comparisons.

This example cannot be performed with 5-year data, as only one set of 5-year data has been released. However, an example using the 3-year data is available in the 2006-2008 ACS Accuracy of the Data document located at:

http://www.census.gov/acs/www/Downloads/data_documentation/Accuracy/accuracy2006-2008ACS3-Year.pdf. This example will be updated when the 2007-2009 data is released.

Control of Nonsampling Error

As mentioned earlier, sample data are subject to nonsampling error. This component of error could introduce serious bias into the data, and the total error could increase dramatically over that which would result purely from sampling. While it is impossible to completely eliminate nonsampling error from a survey operation, the Census Bureau attempts to control the sources of such error during the collection and processing operations. Described below are the primary sources of nonsampling error and the programs instituted for control of this error. The success of these programs, however, is contingent upon how well the instructions were carried out during the survey.

- **Coverage Error** — It is possible for some sample housing units or persons to be missed entirely by the survey (undercoverage), but it is also possible for some sample housing units and persons to be counted more than once (overcoverage). Both the undercoverage and overcoverage of persons and housing units can introduce biases into the data, increase respondent burden and survey costs.

A major way to avoid coverage error in a survey is to ensure that its sampling frame, for ACS an address list in each state, is as complete and accurate as possible. The source of addresses for the ACS is the MAF, which was created by combining the Delivery Sequence File of the United States Postal Service and the address list for Census 2000. An attempt is made to assign all appropriate geographic codes to each MAF address via an automated procedure using the Census Bureau TIGER (Topologically Integrated Geographic Encoding and Referencing) files. A manual coding operation based in the appropriate regional offices is attempted for addresses, which could not be automatically coded. The MAF was used as the source of addresses for selecting sample housing units and mailing questionnaires. TIGER produced the location maps for CAPI assignments. Sometimes the MAF has an address that is the duplicate of another address already on the MAF. This could occur when there is a slight difference in the address such as 123 Main Street versus 123 Maine Street.

In the CATI and CAPI nonresponse follow-up phases, efforts were made to minimize the chances that housing units that were not part of the sample were interviewed in place of units in sample by mistake. If a CATI interviewer called a mail nonresponse case and was not able to reach the exact address, no interview was conducted and the case was eligible for CAPI. During CAPI follow-up, the interviewer had to locate the exact address for each sample housing unit. If the interviewer could not locate the exact sample unit in a multi-unit structure, or found a different number of units than expected, the interviewers were instructed to list the units in the building and follow a specific procedure to select a replacement sample unit. Person overcoverage can occur when an individual is included as a member of a housing unit but does not meet ACS residency rules.

Coverage rates give a measure of undercoverage or overcoverage of persons or housing units in a given geographic area. Rates below 100 percent indicate undercoverage, while rates above 100 percent indicate overcoverage. Coverage rates are released concurrent with the release of estimates on American FactFinder in the B98 series of detailed tables. Further information about ACS coverage rates may be found at http://www.census.gov/acs/www/methodology/coverage_rates_data/.

- Nonresponse Error — Survey nonresponse is a well-known source of nonsampling error. There are two types of nonresponse error – unit nonresponse and item nonresponse. Nonresponse errors affect survey estimates to varying levels depending on amount of nonresponse and the extent to which nonrespondents differ from respondents on the characteristics measured by the survey. The exact amount of nonresponse error or bias on an estimate is almost never known. Therefore, survey researchers generally rely on proxy measures, such as the nonresponse rate, to indicate the potential for nonresponse error.
 - o Unit Nonresponse — Unit nonresponse is the failure to obtain data from housing units in the sample. Unit nonresponse may occur because households are unwilling or unable to participate, or because an interviewer is unable to make contact with a housing unit. Unit nonresponse is problematic when there are systematic or variable differences between interviewed and noninterviewed housing units on the characteristics measured by the survey. Nonresponse bias is introduced into an estimate when differences are systematic, while nonresponse error for an estimate evolves from variable differences between interviewed and noninterviewed households.

The ACS makes every effort to minimize unit nonresponse, and thus, the potential for nonresponse error. First, the ACS used a combination of mail, CATI, and CAPI data collection modes to maximize response. The mail phase included a series of three to four mailings to encourage housing units to return the questionnaire. Subsequently, mail nonrespondents (for which phone numbers are available) were contacted by CATI for an interview. Finally, a subsample of the mail and telephone nonrespondents was contacted by a personal visit to attempt an interview. Combined, these three efforts resulted in a very high overall response rate for the ACS.

ACS response rates measure the percent of units with a completed interview. The higher the response rate, and consequently the lower the nonresponse rate, the less chance estimates may be affected by nonresponse bias. Response and nonresponse rates, as well as rates for specific types of nonresponse, are released concurrent with the release of estimates on American FactFinder in the B98 series of detailed tables. Further information about response and nonresponse rates may be found at http://www.census.gov/acs/www/methodology/response_rates_data/.

- o Item Nonresponse — Nonresponse to particular questions on the survey questionnaire and instrument allows for the introduction of error or bias into the data, since the characteristics of the nonrespondents have not been observed and may differ from those reported by respondents. As a result, any imputation procedure using respondent data may not completely reflect this difference either at the elemental level (individual person or housing unit) or on average.

Some protection against the introduction of large errors or biases is afforded by minimizing nonresponse. In the ACS, item nonresponse for the CATI and CAPI operations was minimized by the requirement that the automated instrument receive a response to each question before the next one could be asked. Questionnaires returned by mail were edited for completeness and acceptability. They were reviewed by computer for content omissions and population coverage. If necessary, a telephone follow-up was made to obtain missing information. Potential coverage errors were included in this follow-up.

Allocation tables provide the weighted estimate of persons or housing units for which a value was imputed, as well as the total estimate of persons or housing units that were eligible to answer the question. The smaller the number of imputed responses, the lower the chance that the item nonresponse is contributing a bias to the estimates. Allocation tables are released concurrent with the release of estimates on American Factfinder in the B99 series of detailed tables with the overall allocation rates across all person and housing unit characteristics in the B98 series of detailed tables. Additional information on item nonresponse and allocations can be found at http://www.census.gov/acs/www/methodology/item_allocation_rates_data/.

- Measurement and Processing Error — The person completing the questionnaire or responding to the questions posed by an interviewer could serve as a source of error, although the questions were cognitively tested for phrasing, and detailed instructions for completing the questionnaire were provided to each household.
 - o Interviewer monitoring — The interviewer may misinterpret or otherwise incorrectly enter information given by a respondent; may fail to collect some of the information for a person or household; or may collect data for households that were not designated as part of the sample. To control these problems, the work of interviewers was monitored carefully. Field staff were prepared for their tasks by using specially developed training packages that included hands-on experience in using survey materials. A sample of the households interviewed by CAPI

interviewers was reinterviewed to control for the possibility that interviewers may have fabricated data.

- Processing Error — The many phases involved in processing the survey data represent potential sources for the introduction of nonsampling error. The processing of the survey questionnaires includes the keying of data from completed questionnaires, automated clerical review, follow-up by telephone, manual coding of write-in responses, and automated data processing. The various field, coding and computer operations undergo a number of quality control checks to insure their accurate application.
- Content Editing — After data collection was completed, any remaining incomplete or inconsistent information was imputed during the final content edit of the collected data. Imputations, or computer assignments of acceptable codes in place of unacceptable entries or blanks, were needed most often when an entry for a given item was missing or when the information reported for a person or housing unit on that item was inconsistent with other information for that same person or housing unit. As in other surveys and previous censuses, the general procedure for changing unacceptable entries was to allocate an entry for a person or housing unit that was consistent with entries for persons or housing units with similar characteristics. Imputing acceptable values in place of blanks or unacceptable entries enhances the usefulness of the data.

Issues with Approximating the Standard Error of Linear Combinations of Multiple Estimates

Several examples are provided here to demonstrate how different the approximated standard errors of sums can be compared to those derived and published with ACS microdata.

- A. With the release of the 5-year data, detailed tables down to tract and block group will be available. At these geographic levels, many estimates may be zero. As mentioned in the ‘Calculations of Standard Errors’ section, a special procedure is used to estimate the MOE when an estimate is zero. For a given geographic level, the MOEs will be identical for zero estimates. When summing estimates which include many zero estimates, the standard error and MOE in general will become unnaturally inflated. Therefore, users are advised to sum only one of the MOEs from all of the zero estimates.

Suppose we wish to estimate the total number of people whose first reported ancestry was ‘Subsaharan African’ in Rutland County, Vermont.

Table A: 2005-2009 Ancestry Categories from Table B04001: First Ancestry Reported

First Ancestry Reported Category	Estimate	MOE
Subsaharan African:	48	43
Cape Verdean	9	15
Ethiopian	0	93
Ghanian	0	93
Kenyan	0	93
Liberian	0	93
Nigerian	0	93
Senegalese	0	93
Sierra Leonean	0	93
Somalian	0	93
South African	10	16
Sudanese	0	93
Ugandan	0	93
Zimbabwean	0	93
African	20	33
Other Subsaharan African	9	16

2009 American FactFinder

To estimate the total number of people, we add up all of the categories.

$$\text{Total Number of People} = 9 + 0 + \dots + 0 + 10 + 0 \dots + 20 + 9 = 48$$

To approximate the standard error using all of the MOEs we obtain:

$$SE(total) = \sqrt{\left(\frac{15}{1.645}\right)^2 + \left(\frac{93}{1.645}\right)^2 + \cdots + \left(\frac{16}{1.645}\right)^2 + \cdots + \left(\frac{33}{1.645}\right)^2 + \left(\frac{16}{1.645}\right)^2} = 189.3$$

Using only one of the MOEs from the zero estimates, we obtain:

$$SE(total) = \sqrt{\left(\frac{15}{1.645}\right)^2 + \left(\frac{93}{1.645}\right)^2 + \left(\frac{16}{1.645}\right)^2 + \left(\frac{33}{1.645}\right)^2 + \left(\frac{16}{1.645}\right)^2} = 62.2$$

From the table, we know that the actual MOE is 43, giving a standard error of $43 / 1.645 = 26.1$. The first method is roughly seven times larger than the actual standard error, while the second method is roughly 2.4 times larger.

Leaving out all of the MOEs from zero estimates we obtain:

$$SE(total) = \sqrt{\left(\frac{15}{1.645}\right)^2 + \left(\frac{16}{1.645}\right)^2 + \left(\frac{33}{1.645}\right)^2 + \left(\frac{16}{1.645}\right)^2} = 26.0$$

In this case, it is very close to the actual SE. This is not always the case, as can be seen in the examples below.

- B. Suppose we wish to estimate the total number of males with income below the poverty level in the past 12 months using both state and PUMA level estimates for the state of Wyoming. Part of the detailed table B17001² is displayed below with estimates and their margins of error in parentheses.

² Table C17001 is used in this example for the 2009 1-year Accuracy documents. C17001 is not published for the 2005-2009 5-year data.

Table B: 2005-2009 ACS estimates of Males with Income Below Poverty from table B17001:
Poverty Status in the Past 12 Months by Sex by Age

Characteristic	Wyoming	PUMA 00100	PUMA 00200	PUMA 00300	PUMA 00400
Male	21,769 (1,480)	4,496 (713)	5,891 (622)	4,706 (665)	6,676 (742)
Under 5 Years	3,064 (422)	550 (236)	882 (222)	746 (196)	886 (237)
5 Years Old	348 (106)	113 (65)	89 (57)	82 (55)	64 (44)
6 to 11 Years Old	2,424 (421)	737 (272)	488 (157)	562 (163)	637 (196)
12 to 14 Years Old	1,281 (282)	419 (157)	406 (141)	229 (106)	227 (111)
15 Years Old	391 (128)	51 (37)	167 (101)	132 (64)	41 (38)
16 and 17 Years Old	779 (258)	309 (197)	220 (91)	112 (72)	138 (112)
18 to 24 Years old	4,504 (581)	488 (192)	843 (224)	521 (343)	2,652 (481)
25 to 34 Years Old	2,289 (366)	516 (231)	566 (158)	542 (178)	665 (207)
35 to 44 Years Old	2,003 (311)	441 (122)	535 (160)	492 (148)	535 (169)
45 to 54 Years Old	1,719 (264)	326 (131)	620 (181)	475 (136)	298 (113)
55 to 64 Years Old	1,766 (323)	343 (139)	653 (180)	420 (135)	350 (125)
65 to 74 Years Old	628 (142)	109 (69)	207 (77)	217 (72)	95 (55)
75 Years and Older	573 (147)	94 (53)	215 (86)	176 (72)	88 (62)

2009 American FactFinder

The first way is to sum the thirteen age groups for Wyoming:

$$\text{Estimate(Male)} = 3,064 + 348 + \dots + 573 = 21,769.$$

The first approximation for the standard error in this case gives us:

$$SE(\text{Male}) = \sqrt{\left(\frac{422}{1.645}\right)^2 + \left(\frac{106}{1.645}\right)^2 + \dots + \left(\frac{147}{1.645}\right)^2} = 696.6$$

A second way is to sum the four PUMA estimates for Male to obtain:

$$\text{Estimate(Male)} = 4,496 + 5,891 + 4,706 + 6,676 = 21,769 \text{ as before.}$$

The second approximation for the standard error yields:

$$SE(\text{Male}) = \sqrt{\left(\frac{713}{1.645}\right)^2 + \left(\frac{622}{1.645}\right)^2 + \left(\frac{665}{1.645}\right)^2 + \left(\frac{742}{1.645}\right)^2} = 835.3$$

Finally, we can sum up all thirteen age groups for all four PUMAs to obtain an estimate based on a total of 52 estimates:

$$Estimate(Male) = 550 + 113 + \dots + 88 = 21,769$$

And the third approximated standard error is

$$SE(Male) = \sqrt{\left(\frac{422}{1.645}\right)^2 + \left(\frac{106}{1.645}\right)^2 + \dots + \left(\frac{62}{1.645}\right)^2} = 721.9$$

However, we do know that the standard error using the published MOE is $1,480 / 1.645 = 899.7$. In this instance, all of the approximations under-estimate the published standard error and should be used with caution.

- C. Suppose we wish to estimate the total number of males at the national level using age and citizenship status. The relevant data from table B05003 is displayed in table C below.

Table C: 2005-2009 ACS estimates of males from B05003: Sex by Age by Citizenship Status

Characteristic	Estimate	MOE
Male	148,535,646	6,574
Under 18 Years	37,971,739	6,285
Native	36,469,916	10,786
Foreign Born	1,501,823	11,083
Naturalized U.S. Citizen	282,744	4,284
Not a U.S. Citizen	1,219,079	10,388
18 Years and Older	110,563,907	6,908
Native	93,306,609	57,285
Foreign Born	17,257,298	52,916
Naturalized U.S. Citizen	7,114,681	20,147
Not a U.S. Citizen	10,142,617	53,041

2009 American FactFinder

The estimate and its MOE are actually published. However, if they were not available in the tables, one way of obtaining them would be to add together the number of males under 18 and over 18 to get:

$$Estimate(Male) = 37,971,739 + 110,563,907 = 148,535,646$$

And the first approximated standard error is

$$SE(Male) = SE(148,535,646) = \sqrt{\left(\frac{6,285}{1.645}\right)^2 + \left(\frac{6,908}{1.645}\right)^2} = 5,677.4$$

Another way would be to add up the estimates for the three subcategories (Native, and the two subcategories for Foreign Born: Naturalized U.S. Citizen, and Not a U.S. Citizen), for males under and over 18 years of age. From these six estimates we obtain:

Estimate(Male)

$$= 36,469,916 + 282,744 + 1,219,079 + 93,306,609 + 7,114,681 \\ + 101,42,617 = 148,535,646$$

With a second approximated standard error of:

$$SE(Male) = SE(148,535,646)$$

$$= \sqrt{\left(\frac{10,786}{1.645}\right)^2 + \left(\frac{4,284}{1.645}\right)^2 + \left(\frac{10,388}{1.645}\right)^2 + \left(\frac{57,285}{1.645}\right)^2 + \left(\frac{20,147}{1.645}\right)^2 + \left(\frac{53,041}{1.645}\right)^2} = 49,920.0$$

We do know that the standard error using the published margin of error is $6,574 / 1.645 = 3,996.4$. With a quick glance, we can see that the ratio of the standard error of the first method to the published-based standard error yields 1.42; an over-estimate of roughly 42%, whereas the second method yields a ratio of 12.49 or an over-estimate of 1,149%. This is an example of what could happen to the approximate SE when the sum involves a controlled estimate. In this case, it is sex by age.

- D. Suppose we are interested in the total number of people aged 65 or older and its standard error. Table D shows some of the estimates for the national level from table B01001 (the estimates in gray were derived for the purpose of this example only).

Table D: Some Estimates from AFF Table B01001: Sex by Age for 2005-2009

Age Category	Estimate, Male	MOE, Male	Estimate, Female	MOE, Female	Total	Approximated MOE, Total
65 and 66 years old	2,248,426	8,047	2,532,831	9,662	4,781,257	12,574
67 to 69 years old	2,834,475	8,953	3,277,067	8,760	6,111,542	12,526
70 to 74 years old	3,924,928	8,937	4,778,305	10,517	8,703,233	13,801
75 to 79 years old	3,178,944	9,162	4,293,987	11,355	7,472,931	14,590
80 to 84 years old	2,226,817	6,799	3,551,245	9,898	5,778,062	12,008
85 years and older	1,613,740	7058	3,540,105	10,920	5,153,845	13,002
Total	16,027,330	20,119	21,973,540	25,037	38,000,870	32,119

2009 American FactFinder

To begin we find the total number of people aged 65 and over by simply adding the totals for males and females to get $16,027,330 + 21,973,540 = 38,000,870$. One way we could use is summing males and female for each age category and then using their MOEs to approximate the standard error for the total number of people over 65.

$$MOE(\text{Age 65 and 66}) = MOE(4,781,257) = \sqrt{8,047^2 + 9,662^2} = 12,574$$

$$MOE(\text{Age 67 to 69}) = MOE(6,111,542) = \sqrt{8,953^2 + 8,760^2} = 12,526$$

... etc. ...

Now, we calculate for the number of people aged 65 or older to be 38,000,870 using the six derived estimates and approximate the standard error:

$$SE(38,000,870) = \sqrt{\left(\frac{7,644}{1.645}\right)^2 + \left(\frac{7,614}{1.645}\right)^2 + \left(\frac{8,390}{1.645}\right)^2 + \left(\frac{8,870}{1.645}\right)^2 + \left(\frac{7,300}{1.645}\right)^2 + \left(\frac{7,904}{1.645}\right)^2} \\ = 32,119$$

For this example the estimate and its MOE are published in table B09017. The total number of people aged 65 or older is 38,000,870 with a margin of error of 4,944. Therefore the published-based standard error is:

$$SE(38,000,870) = 4,944/1.645 = 3,005.$$

The approximated standard error, using six derived age group estimates, yields an approximated standard error roughly 10.7 times larger than the published-based standard error.

As a note, there are two additional ways to approximate the standard error of people aged 65 and over in addition to the way used above. The first is to find the published MOEs for the males age 65 and older and of females aged 65 and older separately and then combine to find the approximate standard error for the total. The second is to use all twelve of the published estimates together, that is, all estimates from the male age categories and female age categories, to create the SE for people aged 65 and older. However, in this particular example, the results from all three ways are the same. So no matter which way you use, you will obtain the same approximation for the SE. This is different from the results seen in example B.

- E. For an alternative to approximating the standard error for people 65 years and older seen in part D, we could find the estimate and its SE by summing all of the estimate for the ages less than 65 years old and subtracting them from the estimate for the total population. Due to the large number of estimates, Table E does not show all of the age groups. In addition, the estimates in part of the table shaded gray were derived for the purposes of this example only and cannot be found in base table B01001.

Table E: Some Estimates from AFF Table B01001: Sex by Age for 2005-2009:

Age Category	Estimate, Male	MOE, Male	Estimate, Female	MOE, Female	Total	Estimated MOE, Total
Total Population	148,535,646	6,574	152,925,887	6,584	301,461,533	9,304
Under 5 years	10,663,983	3,725	10,196,361	3,557	20,860,344	5,151
5 to 9 years old	10,137,130	15,577	9,726,229	16,323	19,863,359	22,563
10 to 14 years old	10,567,932	16,183	10,022,963	17,199	20,590,895	23,616
...		
62 to 64 years old	3,888,274	11,186	4,257,076	11,970	8,145,350	16,383
Total for Age 0 to 64 years old	132,508,316	48,688	130,952,347	49,105	263,460,663	69,151
Total for Age 65 years and older	16,027,330	49,130	21,973,540	49,544	38,000,870	69,774

2009 American FactFinder

An estimate for the number of people age 65 and older is equal to the total population minus the population between the ages of zero and 64 years old:

Number of people aged 65 and older: $301,461,533 - 263,460,663 = 38,000,870$.

The way to approximate the SE is the same as in part D. First we will sum male and female estimates across each age category and then approximate the MOEs. We will use that information to approximate the standard error for our estimate of interest:

$$MOE(\text{Total Population}) = MOE(301,461,533) = \sqrt{6,574^2 + 6,584^2} = 9,304$$

$$MOE(\text{Under 5 years}) = MOE(20,860,344) = \sqrt{3,725^2 + 3,557^2} = 5,151$$

... etc. ...

And the SE for the total number of people aged 65 and older is:

$$\begin{aligned}
SE(\text{Age 65 and older}) &= SE(38,000,870) \\
&= \sqrt{\left(\frac{9,304}{1.645}\right)^2 + \left(\frac{5,151}{1.645}\right)^2 + \left(\frac{22,563}{1.645}\right)^2 + \left(\frac{23,616}{1.645}\right)^2 + \dots + \left(\frac{16,383}{1.645}\right)^2} \\
&= 42,416
\end{aligned}$$

Again, as in Example D, the estimate and its MOE are published in B09017. The total number of people aged 65 or older is 38,000,870 with a margin of error of 4,944. Therefore the standard error is:

$$SE(38,000,870) = 4,944 / 1.645 = 3,005.$$

The approximated standard error using the seventeen derived age group estimates yields a standard error roughly 14.1 times larger than the actual SE.

Data users can mitigate the problems shown in examples A through E to some extent by utilizing a collapsed version of a detailed table (if it is available) which will reduce the number of estimates used in the approximation. These issues may also be avoided by creating estimates and SEs using the Public Use Microdata Sample (PUMS) or by requesting a custom tabulation, a fee-based service offered under certain conditions by the Census Bureau. More information regarding custom tabulations may be found at http://www.census.gov/acs/www/data_documentation/custom_tabulations/.

Things That May Affect Estimates from the American Community Survey

U S C E N S U S B U R E A U

1

The purpose of this presentation is to provide information on issues than can affect estimates created from American Community Survey data, in particular, sampling error.

Overview

- Definition of sampling error
- Measures associated with sampling error
 - Standard error
 - Margin of error
 - Confidence intervals
 - Coefficient of variation
- How to use measures associated with sampling error
- Non-sampling error
- Population controls

U S C E N S U S B U R E A U

2

First we'll define sampling error, and then we'll discuss four different measures associated with sampling error, including how they are calculated and what they mean.

Then we'll talk about why measures of sampling error are important and how you can use them to help draw appropriate conclusions about ACS data.

Finally we'll talk briefly about non-sampling error and population controls.

Definition of Sampling Error

USCENSUSBUREAU

3

What is Sampling Error?

Definition

The **uncertainty** associated with an estimate that is based on data gathered from a sample of the population rather than the full population

USCENSUSBUREAU

4

What is Sampling Error?

Let's start our discussion with the definition:

- Sampling error is the uncertainty associated with an estimate that is based on data gathered from a **sample** of the population rather than the **full** population.
- So, why do sample estimates have uncertainty associated with them?
There are two reasons:
 1. Estimates of characteristics from the sample data can differ from those that would be obtained if the entire population were surveyed.
 2. Estimates from one subset or sample of the population can differ from those based on a different sample from that same population.

Illustration of Sampling Error

Estimate average number of children per household for a population with 3 households:

Household A	1 child
Household B	2 children
Household C	3 children

Average based on the full population is two children per household

U S C E N S U S B U R E A U

5

Let's illustrate this with a simple example.

Suppose we want to estimate the average number of children per household for a population that only has 3 households. If we use the information for all households in the population, then we would divide the total number of children, six, by the total number of households, three, to get an average of two children per household.

But, suppose we did not have enough money to collect data from all three households, the full population, and instead had to estimate the average by collecting data from only two of the three households.

Conceptualizing Sampling Error

Three different samples:

1. Households A and B (1 child, 2 children)
2. Households B and C (2 children, 3 children)
3. Households A and C (1 child, 3 children)

Three different averages:

1. 1.5 children $(1 + 2) / 2$
2. 2.5 children $(2 + 3) / 2$
3. 2 children $(1 + 3) / 2$

U S C E N S U S B U R E A U

6

There are three different samples of two households that we could potentially select, as shown here. If we randomly chose two households, we could get households A and B with 1 child and 2 children respectively, households B and C with 2 children and 3 children respectively, or households A and C with 1 child and 3 children respectively. If we sum the total number of children in each sample and divide by the total number of households in each sample, we will get the average number of children per household in each sample.

Sample 1 with households A and B has an average of 1.5 children per household (1 child + 2 children equals 3 total children in the sample, divided by 2 total households in the sample), while Sample 2 has an average of 2.5 children per household, and Sample 3 has an average of 2 children per household. So, these three different samples provide three different estimates of the average number of children per household for the **full** population of three households. And, the estimate from Sample 3 is the only one that is identical to the average for the full population of three households that we calculated in the last slide.

The variation between these sample estimates and the actual value for the full population is what we call sampling error. It results from collecting information from only a sample rather than the full population.

Measures associated with sampling error

USCENSUSBUREAU

7

Measures Associated with Sampling Error

- Standard Error (SE)
- Margin of Error (MOE)
- Confidence Interval (CI)
- Coefficient of Variation (CV)

U S C E N S U S B U R E A U

8

There are four key measures of sampling error that ACS data users should understand –standard error, margin of error, confidence interval, and coefficient of variation. Each of these measures are related to one another.

In the next section of this presentation, we will examine each of these measures in detail to understand how each is calculated and interpreted, and to understand how these different measures are related to each other.

Standard Error (SE)

Definition

A measure of the variability of an estimate due to sampling

Depends on variability in the population and sample size

Foundational measure

USCENSUSBUREAU

9

The first measure of sampling error we will discuss is the standard error. The standard error is a measure of the variability of an estimate due to sampling. It indicates the extent to which an estimate derived from a sample survey can be expected to deviate from the population value.

The standard error for an estimate depends upon the underlying variability in the population for the characteristic as well as the sample size used for the survey. For example, if 80 percent of households in a population have two children, then the standard error of the estimate of average children per household will be smaller than in another population where there is more variation among households in the number of children.

The standard error is a foundational measure from which other sampling error measures are derived, and it is required in order to conduct tests of statistical significance. However, standard errors are not usually used for interpretation.

Standard Error (SE)

Formula

$$SE = MOE / 1.645$$

2007 ACS Data for Baltimore City:

52.1% Percent of males who have never married
1.7% Margin of Error

$$SE = 1.7\% / 1.645$$

$$SE = 1.033\%$$

US CENSUS BUREAU

10

The Census Bureau provides the Margin of Error for each published ACS estimate. We'll look at margins of error in more detail a little later, but as this slide shows, the standard error for an ACS estimate can be obtained by dividing the published margin of error for the estimate by the value 1.645. For estimates from years 2005 or earlier, use the value 1.65 with the published margin of error.

Let's calculate a standard error using some of the 2007 ACS data for the city of Baltimore, Maryland. The 2007 ACS provides an estimate of 52.1 percent for males age 15 and older who live in Baltimore city and have never married. The published margin of error for this estimate is 1.7 percent. When we divide the margin of error of 1.7 percent by the value 1.645, we get a standard error of 1.033 percent.

Margin of Error (MOE)

Definition

A measure of the precision of an estimate at a given level of confidence (90%, 95%, 99%)

Confidence level of a MOE

MOEs at the 90% confidence level for all published ACS estimates

US CENSUS BUREAU

11

The margin of error is defined as a measure of the precision of an estimate at a given level of confidence. The most commonly used confidence levels are 90%, 95% and 99%. What does the confidence level of a margin of error mean?

The confidence level of a margin of error indicates the likelihood that the difference between the population value and the sample estimate is less than or equal to the margin of error.

All ACS estimates are published with their margins of error at the 90 percent confidence level. However, it is possible to construct margins of error with higher levels of confidence, such as 95 percent or 99 percent. This is done by adjusting the published margin of error. Instructions for these adjustments can be found in the technical appendices of the The ACS Compass Products Handbooks now available on the American Community Survey web site.

Margin of Error (MOE)

Formula

$$\text{MOE} = \pm 1.645 \times \text{SE (90\% level)}$$

Values for other confidence levels

$$95\% = 1.960$$

$$99\% = 2.576$$

U S C E N S U S B U R E A U

12

Even though all ACS estimates are published with margins of error at the 90 percent confidence level, it's important to understand how that measure is constructed. It will help you understand the relationship between the different measures associated with sampling error.

The margin of error is a multiple of the Standard Error (SE). The margins of error published in the ACS are calculated at the 90 percent confidence level by multiplying the Standard Error by the value 1.645. The 1.645 value is related to the 90 percent confidence level. Any other confidence level, whether 95 percent or 99 percent, will have different values associated with them. Those values will increase as we increase our level of certainty or confidence level.

Interpreting Margin of Error

- Indicates that a data user can be 90 percent certain that the estimate and the population value differ by no more than the value of the MOE
- MOE can help data users assess the reliability of an estimate
- MOE can help data users avoid misinterpreting small differences between estimates as significant

U S C E N S U S B U R E A U

13

How do we interpret the margin of error for an estimate?

At a 90 percent confidence level, the margin of error indicates that there is a 90 percent probability that the estimate and the population value differ by no more than the value of the margin of error. In other words, we can be 90 percent certain that the range established by the margin of error contains the population value.

Margins of error are useful in assessing the reliability of estimates and whether differences between estimates are significant.

Later, we'll talk more about the specific meaning of the word "significant."

Interpreting Margin of Error

Example for Baltimore City:

52.1% Percent of males who have never married
1.7% Margin of Error

- Indicates 90 percent chance that the estimate of 52.1% and the population value differ by no more than 1.7%
- Size of MOE relative to size of estimate

U S C E N S U S B U R E A U

14

Let's continue using our previous example about never married males living in Baltimore, Maryland. The ACS estimate shows that, of the males age 15 and over who live in Baltimore city, 52.1 percent have never married. That estimate has a margin of error of plus or minus 1.7 percent, which establishes a range of 50.4 percent to 53.8 percent. Now, let's say we actually asked every male age 15 or older who lives in Baltimore city if they have ever been married. The margin of error of the ACS estimate means that there is a 90 percent chance that the range of 50.4 percent to 53.8 percent would contain the population value.

In general, the larger the margin of error relative to the size of the estimate, the less reliable is the estimate. The margin of error provides a concise measure of the precision of an ACS sample estimate.

Confidence Interval

Definition

A range that is expected to contain the population value of the characteristic with a known probability.

USCENSUSBUREAU

15

The third key measure of sampling error is the confidence interval. It is defined as a range that is expected to contain the population value of the characteristic with a known probability.

Confidence Interval

Formula

$$L_{CL} = \hat{X} - MOE_{CL}$$

$$U_{CL} = \hat{X} + MOE_{CL}$$

where

L_{CL} is the lower bound at the desired confidence level

U_{CL} is the upper bound at the desired confidence level

\hat{X} is the ACS estimate and

MOE_{CL} is the margin of error at the desired confidence level

U S C E N S U S B U R E A U

16

The confidence interval is calculated by subtracting the margin of error from the estimate, to create the lower bound, and by adding the margin of error to the estimate to create the upper bound. The range between the upper and lower bounds of the confidence interval is expected to contain the population value with a given level of confidence.

Now, let's calculate the confidence interval for the ACS estimate of the percent of never married males in Baltimore city.

Confidence Interval

Calculation Example for Baltimore City

$$52.1\% - 1.7\% = 50.4\%$$

$$52.1\% + 1.7\% = 53.8\%$$

Confidence Interval = 50.4% to 53.8%

USCENSUSBUREAU

17

When we subtract the margin of error from the estimate, we get a lower bound of 50.4 percent, and when we add the margin of error, we get an upper bound of 53.8 percent.

The resulting confidence interval ranges from a low of 50.4 percent of the never married males to a high of 53.8 percent of never married males.

How should this confidence interval be interpreted?

Confidence Interval Interpretation

- We can be 90 percent certain that the confidence interval from 50.4% to 53.8% contains the population value of never married males 15 years and older in Baltimore City
- Useful to display confidence intervals

U S C E N S U S B U R E A U

18

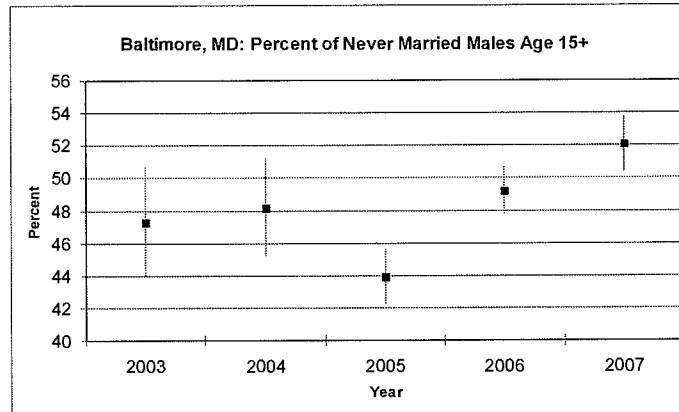
In simple terms, this confidence interval means that we can be 90 percent certain that the confidence interval from 50.4 percent to 53.8 percent contains the population value of never married males 15 years and older in Baltimore City.

If you'll recall during our discussion of margin of error, we said the margin of error establishes a range. The upper and lower bounds of that range define the confidence interval.

In general, for similar size estimates, larger confidence intervals indicate that estimates are less precise and reliable.

Graphing confidence intervals is a useful way to provide a quick visual comparison of the extent of sampling error for different estimates.

Displaying Confidence Intervals



USCENSUSBUREAU

19

In this graph, the dots show the ACS estimate and the lines show the confidence interval. It is easy to see that the confidence intervals are much larger for the 2003 and 2004 ACS estimates for Baltimore.

The confidence intervals for the estimates for 2005 through 2007 are much smaller, indicating that these estimates have smaller margins of error and are more reliable. This makes sense because the sample size of the ACS tripled in 2005 when the survey moved out of its testing phase and expanded to include each county in the United States and Puerto Rico in its annual sample.

Coefficient of Variation (CV)

Definition

The relative amount of sampling error associated with a sample estimate

Formula

$$CV = SE / \text{Estimate} * 100\%$$

USCENSUSBUREAU

20

The fourth measure of sampling error we want to cover in this presentation is the coefficient of variation, or CV. It is defined as the **relative** amount of sampling error associated with a sample estimate, and is used to assess the reliability of an estimate.

The CV is calculated as the ratio of the standard error for an estimate to the estimate itself, usually expressed as a percent, which is why it is multiplied by 100 percent in the formula. A small CV indicates that the sampling error is small relative to the estimate, and therefore the user can be more confident that the estimate is a good approximation to the population value.

Because they are expressed as percents, it is easier to compare the reliability of two different estimates using CVs than it is using margins of error.

Coefficient of Variation (CV)

Example for Baltimore City

Estimate = 52.1% of never married males

Standard Error = 1.033%

$$CV = SE / Estimate * 100\%$$

$$CV = 1.033\% / 52.1\% * 100\%$$

$$CV = 1.98\%$$

U S C E N S U S B U R E A U

21

Now, continuing our example, let's calculate the CV for the ACS estimate of males age 15 and over living in Baltimore who have never married.

The estimate for never married males is 52.1%, and we calculated its standard error as 1.033%. The CV equals the standard error divided by the estimate, and we get a value for this CV of 1.98%.

So, how should CVs be interpreted?

Interpreting Coefficients of Variation

- Size of the CV
- In Baltimore City example, the CV is small ($< 2\%$) indicating this is a reliable estimate
- No hard-and-fast rules about the size of CVs
- Caution for proportions close to zero

USCENSUSBUREAU

22

The size of a CV is important for its interpretation. In general, the smaller the CV, the more reliable the estimate.

In the sample calculation we just completed for Baltimore city, we found a CV of only 1.98 percent. In other words, the sampling error for this estimate is about 2 percent the size of the estimate.

However, there are no hard and fast rules about the usability of an estimate based on the size of the CV. Smaller CVs imply better estimates, but the question of “when is the CV too large?” cannot be answered in a vacuum. It all depends on the context and how the estimate will be used. For example, if you just want to know if an estimate is below a certain value, an estimate with a large CV may be reliable enough for this purpose. Someone else may need an extremely precise estimate, and even a small CV (as small as 2 percent) may be too large for their use.

For proportions close to zero, the CV may be unstable. In this case, data users are advised to rely on the margin of error or confidence interval rather than the CV.

In general, data users are encouraged to consider both the margin of error and the CV when evaluating the reliability of an estimate.

Sampling Error is Related to Sample Size

- The larger the sample size, the smaller the uncertainty or sampling error
- Combining ACS data from multiple years increases sample size and reduces sampling error
- All sample surveys have sampling error – including decennial census long-form data

U S C E N S U S B U R E A U

23

Sampling error is related to sample size. Specifically, the larger the sample size, the smaller the uncertainty or sampling error for estimates based on the survey sample.

Recall the example of estimating the average number of children per household. Now imagine instead of three households we had a full population of 100 households. It's easy to see that our estimate of average children per household generally would have much more uncertainty if we sampled only two of those 100 households, rather than say 30 of those 100 households.

This relationship between sample size and sampling error is the reason that ACS estimates for smaller geographic areas are based on multiple years of data. Combining ACS data from multiple years increases the sample size and in turn reduces the uncertainty or error associated with ACS estimates.

Sampling error is not unique to the ACS – all sample surveys have it, including the decennial census long form. Because the long form was collected from only a sample of all U.S. households, long-form estimates also have sampling error associated with them. Instructions for calculating standard errors for long form estimates can be found in the technical documentation for Summary File 3 or SF-3 on the Census Bureau's website.

How to use measures associated with sampling error

USCENSUSBUREAU

24

How are Measures of Sampling Error Used?

- To indicate the statistical reliability and usability of estimates
- To make comparisons between estimates
- To conduct tests of statistical significance
- To help users draw appropriate conclusions about data

USCENSUSBUREAU

25

So far, we have examined only one of the important ways measures of sampling error can be used – to determine whether an estimate is reliable and usable.

There are two additional ways sampling error measures can be used:

1. To make comparisons between two or more estimates, for example to conduct formal tests of statistical significance, and
2. To help data users draw appropriate conclusions about the data.

The case study we'll examine in the next section of the presentation shows how to use measures of sampling error in each of these ways.

Case Study

Tracking Economic Well-Being in Washington, DC

- In 2005, city implements a series of job training initiatives to increase employment and reduce poverty rates
- In 2008, public officials want to assess changes in poverty rates in the city

U S C E N S U S B U R E A U

26

Suppose that public officials in Washington, DC have implemented a series of job training initiatives they hope will increase employment and reduce poverty rates.

In 2008, after the programs have been in place for several years, officials want to assess whether there has been any change in poverty rates in the city.

Finding the Data

- Washington, DC has a population size greater than 65,000
- Comparable data for both 2006 and 2007 are available from the ACS
- Examine change in the percent of people living in poverty from 2006 to 2007

USCENSUSBUREAU

27

Where can Washington, DC officials find data to address this question?

Since the resident population in Washington, DC exceeds 65,000, there are comparable data for both 2006 and 2007 from the ACS available in American FactFinder.

The 2006 ACS provides a good starting point because it provides some information about the economic characteristics of 2005. This is due to the fact that income data collected in the 2006 ACS are based on income in the “previous 12 months.”

Officials can use the ACS to examine change in the percent of people who were living in poverty in 2006 versus 2007.

Finding the Data

2006 ACS data for Washington, DC

19.6% % of all people living in poverty

1.4% Margin of error

2007 ACS data for Washington, DC

16.4% % of all people living in poverty

1.4% Margin of error

US CENSUS BUREAU

28

City data analysts can use the Census Bureau's American FactFinder website to find the estimates of people living in poverty, along with the margins of error for each estimate.

In 2006, the ACS found 19.6% of all people living in poverty in Washington DC, with a margin of error of 1.4%. The 2007 estimate for percent of people living in poverty was 16.4%, with a margin of error of 1.4%.

These estimates appear to indicate that the poverty rate has declined between 2006 and 2007. But, the analysts know they must first carefully evaluate several measures of sampling error to determine whether the estimates are reliable and usable.

Are the Estimates Reliable and Usable?

Check CVs for each estimate

$$2006: SE = 0.85\% = (1.4\% / 1.645)$$

$$\underline{CV = 4.3\%} = (0.85\% / 19.6\%) * 100$$

$$2007: SE = 0.85\% = (1.4\% / 1.645)$$

$$\underline{CV = 5.2\%} = (0.85\% / 16.4\%) * 100$$

Result = Both estimates are reliable

U S C E N S U S B U R E A U

29

The analysts first calculate standard errors for each of the estimates using the margins of error provided in the AFF tables. They use the denominator of 1.645 to calculate the standard errors for both estimates.

The standard error for the 2006 estimate is equal to the margin of error, 1.4%, divided by 1.645, which equals 0.85%. The CV is the standard error, 0.85%, divided by the estimate, 19.6%, which yields a CV of 4.3%. Similar calculations for 2007 give a CV of 5.2%. Both CVs are relatively small – about 5 percent in both cases.

The analysts also note that neither estimate of the percent of the population living in poverty is close to zero, so the CVs are appropriate measures of reliability.

The analysts conclude that the estimates for both 2006 and 2007 are reliable and usable for their analysis.

Comparing the Estimates

Compare Confidence Intervals:

2006: 18.2% - 21.0% (19.6 +/- 1.4)

2007: 15.0% - 17.8% (16.4 +/- 1.4)

- Is there a significant difference?

USCENSUSBUREAU

30

The analysts begin their comparison of these two estimates by constructing the confidence intervals around each estimate.

The analysts decide they must conduct a formal test of statistical significance.

Test of Statistical Significance

Definition

A test to determine if it is unlikely that something has occurred by chance

A “statistically significant difference” means there is statistical evidence that there is a difference

USCENSUSBUREAU

31

When comparing survey estimates, it is very important that the comparison takes into account the sampling error associated with each estimate.

A test of statistical significance provides **statistical** evidence that indicates whether an observed difference between two estimates is likely due to chance, or likely represents a true difference that exists in the population as a whole.

Conducting Tests of Statistical Significance

Formula

$$\left| \frac{\hat{X}_1 - \hat{X}_2}{\sqrt{SE_1^2 + SE_2^2}} \right| > Z_{CL}$$

where Z_{CL} is the critical value for the desired confidence level

Z_{CL} for 90% confidence level = 1.645

USCENSUSBUREAU

32

This slide shows the formula to calculate a test of statistical significance between two estimates. In the formula, X1 and X2 represent the two estimates, and SE1 is the standard error for X1 and SE2 is the standard error for X2.

Z is the critical value for the desired confidence level for the test. For a 90% confidence level, Z is equal to 1.645.

Testing for Statistical Significance

Substituting the appropriate values:

$$\left| \frac{19.6 - 16.4}{\sqrt{(0.85)^2 + (0.85)^2}} \right| = 2.662$$

- 2.662 \gg 1.645
- Difference is statistically significant at the 90% confidence level

USCENSUSBUREAU

33

The city data analysts plug in the appropriate values to determine whether the decline in the poverty rate in Washington, DC between 2006 and 2007 is statistically significant. They choose a desired confidence level of 90 percent. As indicated in the previous slide, the critical value for a 90 percent confidence level is 1.645.

Their calculation yields an absolute value of 2.662, which is indeed greater than 1.645. They conclude that the decline between 2006 and 2007 in the percent of people living in poverty is statistically significant at the 90 percent confidence level.

Drawing Appropriate Conclusions

- Short-term fluctuations versus real trends
- Increasing confidence level to 95% or 99%
 - Z_{CL} for 95% confidence level = 1.960
 - Z_{CL} for 99% confidence level = 2.576

USCENSUSBUREAU

34

Although the data analysts report back to Washington, DC public officials that the decline between 2006 and 2007 was statistically significant at the 90 percent confidence level, they caution that it is too soon to tell if this is a real trend, or just a short-term fluctuation in the poverty rates. The analysts recommend that public officials track poverty rates for several more years to determine if the 2006 to 2007 decline continues.

The critical value for the test at the 95 percent confidence level is 1.96, and the critical value for a test at the 99 percent confidence level is 2.576. It can be noted that the difference between 2006 and 2007 is statistically significant at both the 95 and 99 percent confidence levels, as well.

Non-sampling error

USCENSUSBUREAU

35

What is Non-Sampling Error?

Definition

Any error affecting a survey or census estimate apart from sampling error

Occurs in complete censuses as well as in sample surveys

U S C E N S U S B U R E A U

36

It is also important to mention that there is an additional source of error in all censuses and surveys, including the ACS. It is called non-sampling error.

Non-sampling error refers to any error associated with a survey or census estimate other than sampling error. The important distinction is that, unlike sampling error, non-sampling error occurs in both sample surveys and censuses.

Types of Non-Sampling Error

- Non-Response Error
- Response Error
- Processing Error
- Coverage Error

USCENSUSBUREAU

37

There are four major types of non-sampling error – non-response error, response error, processing error, and coverage error.

Non-response error occurs when some households or people refuse to participate at all in a census or survey, or when they refuse to answer some questions (which is called item non-response). If the total or item non-response rates are high, then the estimates might be biased. This will happen if the characteristics of those who don't respond are different from the characteristics of those who do respond.

Response error occurs because some respondents may intentionally or unintentionally misreport information. For example, they might not accurately remember what their income was in the last twelve months, or they may not want to report previously undisclosed income.

The third major type of non-sampling error is processing error. This could occur during data entry or coding of survey responses.

Finally, coverage error occurs when a housing unit or a person does not have a chance of being selected for a census or survey. For example, if the census or the ACS did not include newly constructed housing units, then the people who lived in those housing units would be excluded, and this could cause coverage error in the data. In addition, coverage error can occur if a person is incorrectly included in or omitted from a household in sample.

Non-sampling error is difficult to measure directly but there are indirect measures available on the Census Bureau's website.

Population controls

USCENSUSBUREAU

38

One way to reduce coverage error is the use of population controls.

Population Controls

- Independent information used to increase the precision of the ACS estimates
- Reduces sampling and non-sampling errors in the ACS estimates
- Time series of population estimates are revised annually but the ACS estimates for previous years are not.

U S C E N S U S B U R E A U

39

Independent estimates from the Census Bureau's Population Estimate Program are used as controls for the ACS. Housing unit estimates are controlled at the county level for total housing, while total population estimates are controlled at the county level (or groups of counties) by age, sex, race, and Hispanic origin. The group quarters total population is controlled at the state level by major type of group quarters.

However, ACS estimates for these characteristics will not necessarily match the independent estimates. Total population for counties and states will generally match, but other estimates, such as total housing units, population by race, and total population for areas below counties, generally will not match, although they will likely be close.

The use of population controls is important to reduce sampling and non-sampling error in the ACS estimates. In particular, using controls reduces coverage error due to both missing housing units, and also missing people within housing units or group quarters.

As more current data become available, the time series of population estimates back to the preceding census are revised, while the ACS estimates for previous years are not. Therefore, some difference in the ACS estimates across time may be due to changes between the revised series.

The ACS is designed to measure the characteristics of the population and housing units. The official estimates of basic population and housing unit totals still come from the decennial census and from the Census Bureau's Population Estimates Program.

Summary

USCENSUSBUREAU

40

What Have We Learned?

- All surveys have sampling and non-sampling error
- Four key measures of sampling error are standard error, margin of error, confidence interval, and coefficient of variation
- Measures of sampling error provide important information about the reliability of ACS estimates

U S C E N S U S B U R E A U

41

Let's recap the main topics we have covered in this presentation.

First, all surveys have sampling error and non-sampling error, including the ACS and the decennial census long form. Sampling error results from collecting information from only a sample rather than the full population. Non-sampling error results from issues with response, processing, and coverage and is difficult to measure directly.

There are four key measures of sampling error that ACS data users should understand. They are standard error, margin of error, confidence interval, and coefficient of variation. The Census Bureau provides the MOE for most published ACS estimates, and data users can derive the standard error, confidence interval, and coefficient of variation from the MOE.

These four measures of sampling error provide important information about the reliability of ACS estimates. Data users should not use ACS estimates without first assessing their reliability.

What Have We Learned?

- Sampling error measures can be used to make comparisons between estimates and to conduct tests of statistical significance
- Understanding and using measures of sampling error helps users draw appropriate conclusions about ACS data

U S C E N S U S B U R E A U

42

In addition to providing information about the reliability of ACS estimates, sampling error measures can also be used to make comparisons between estimates and to conduct tests of statistical significance.

Finally, it is only by understanding and correctly using measures of sampling error that data users can draw appropriate conclusions about ACS data.

For more information

Subscribe to "ACS Alert"

<http://www.census.gov/acs/www/Special/Alerts.htm>

Visit the ACS/PRCS website:

<http://www.census.gov/acs/www>

Contact by telephone:

1-800-923-8282

Contact by email:

acso.users.support@census.gov

U S C E N S U S B U R E A U

43

This presentation gave you an overview of sampling error and other things that may affect the estimates from the American Community Survey.

The American Community Survey staff has developed the ACS Alert, which is an e-mail newsletter giving data users the latest news about the survey. You can subscribe to the newsletter by contacting the American Community Survey staff or read past editions of the "ACS Alert" on the Internet at:

<http://www.census.gov/acs/www/Special/Alerts.htm>

Please feel free to contact the Census Bureau if you have questions or need further information. If you have questions that are not answered by the Web site, please call 1-800-923-8282 or email acso.users.support@census.gov.

Potter, Andrew

From: CSG, Redistricting 2011
Sent: Thursday, April 07, 2011 5:23 PM
To: Adam Day; Andrea Skorepa; Deanna Weeks; Dennis Ridz; Michel Anderson
Cc: Pettingill, William L; Potter, Andrew
Subject: Redistricting Information Flyer

Redistricting Advisory Committee,

For your information, following is a link to a general flyer about County redistricting that includes a schedule of upcoming meetings of the Redistricting Advisory Committee. This material may be useful for you to share in the community, in addition to the public outreach efforts by the County.

http://www.sdcounty.ca.gov/redistricting/Resource_documents/RedistrictingInfoFlier.pdf. Soon the flyer will be posted in Filipino, Spanish and Vietnamese.

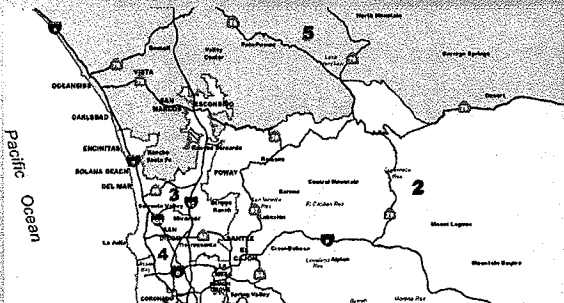
Please let me know if I can answer any questions.

Nicole

Nicole Cretelle Temple, MBA
Redistricting Project Manager
County of San Diego
(619) 531-5276
redistricting2011@sdcounty.ca.gov



Please consider the environment before printing this e-mail



San Diego County's 2011 REDISTRICTING

San Diego County's 2011 Redistricting

You are invited to participate in the County's 2011 redistricting process. A series of upcoming public meetings and the County's redistricting website are two good ways to learn more and get involved.

What is County redistricting?

State law requires the County Board of Supervisors to adjust the boundaries of the County's five supervisorial districts every ten years to balance their populations based on the federal census.

These adjustments may change which supervisorial district residents live in, so the public has a stake.

The Board has appointed a Redistricting Advisory Committee to help gather public opinion through a series of meetings.

After gathering this input, the Committee will prepare a report to the Board of Supervisors that will include up to three recommended redistricting plans. The Board will gather more public comment, then introduce and adopt a redistricting plan that will be effective in early September.

Connect with Redistricting

Public participation is a key part of redistricting, and we want to hear from you! Visit the County's redistricting website to learn more about the process and how to share your ideas.

The website is a one-stop resource, with answers to frequently-asked questions and other helpful information.

Video of all Redistricting Advisory Committee meetings streams live on the site, and the public can view past meetings there too.

Members of the public who want to submit a full redistricting plan or give a suggestion to the Redistricting Advisory Committee should visit the website for more information. The deadline for submitting plans is May 9.

The site is continually updated, and residents can subscribe on the website to receive email alerts when there is new content posted. All Redistricting Advisory Committee meetings, agendas, minutes and proposed redistricting plans will be posted to the website.

www.sdcountry.ca.gov/redistricting

UPCOMING REDISTRICTING ADVISORY COMMITTEE MEETINGS

Members of the public are encouraged to attend and share their ideas

April 11, 9:00 a.m.

Room 310, County Administration Center
1600 Pacific Highway, San Diego, 92101

April 20, 6:00 p.m.

Chula Vista Council Chambers
276 Fourth Avenue, Chula Vista, 91910

April 25, 6:00 p.m.

La Mesa City Council Chambers
8130 Allison Avenue, La Mesa, 91942

April 27, 6:00 p.m.

San Marcos City Council Chambers
1 Civic Center Drive, San Marcos, 92069

April 28, 6:00 p.m.

Solana Beach Council Chambers
635 S. Highway 101, Solana Beach, CA 92075

May 3, 6:00 p.m.

County Operations Center Annex
Planning Commission Hearing Room
5201 Ruffin Road, San Diego, 92123

May 9, 9:00 a.m.

Room 310, County Administration Center
1600 Pacific Highway, San Diego, 92101

May 16, 9:00 a.m.

Room 310, County Administration Center
1600 Pacific Highway, San Diego, 92101

June 2, 2:00 p.m.

Room 310, County Administration Center
1600 Pacific Highway, San Diego, 92101

June 9, 2:00 p.m.

Room 310, County Administration Center
1600 Pacific Highway, San Diego, 92101

June 13, 9:00 a.m.

Room 310, County Administration Center
1600 Pacific Highway, San Diego, 92101



Potter, Andrew

From: CSG, Redistricting 2011
Sent: Friday, April 08, 2011 3:16 PM
To: Abbi Lawrance; CSG, Redistricting 2011
Cc: Potter, Andrew
Subject: RE: Redistricting

Dear Ms. Lawrance,

Thank you for your e-mail about keeping each city in the same district. Your comments will be shared with the Redistricting Advisory Committee. For your reference, the County's 2011 Redistricting Criteria and Guidelines state, "With the exception of the City of San Diego and to the extent possible, individual cities will not be divided between districts."

Sincerely,

Nicole Cretelle Temple, MBA
Redistricting Project Manager
County of San Diego
(619) 531-5276
redistricting2011@sdcounty.ca.gov



Please consider the environment before printing this e-mail

From: Abbi Lawrance [<mailto:abbilawrance@cox.net>]
Sent: Friday, April 08, 2011 10:59 AM
To: CSG, Redistricting 2011
Subject: Redistricting

Please keep each city in the same district if possible. I identify with Encinitas and don't want it split up.
Thank you,
Abbi Lawrance

Potter, Andrew

From: CSG, Redistricting 2011
Sent: Wednesday, April 13, 2011 5:02 PM
To: CSG, Redistricting 2011
Subject: COUNTY REDISTRICTING 2011: IT'S TIME TO BE HEARD

We want to hear from you! Over the next few weeks, the County of San Diego's Redistricting Advisory Committee will convene in each of the County's five districts in a series of public meetings, giving residents an opportunity to learn about County redistricting and voice your ideas.

At each meeting, County staff will present an overview of the County's redistricting process, including the laws, principles and timelines that guide it. The presentation will review population and demographic data and will explain how the public can submit redistricting plans, give suggestions and stay involved using the County's redistricting website.

The meetings are also opportunities for residents to address the Redistricting Advisory Committee with suggestions, plans and information about your community.

District Meeting Schedule

- **April 20, 6 p.m. Chula Vista Council Chambers**, 276 Fourth Avenue, Chula Vista
- **April 25, 6 p.m. La Mesa City Council Chambers** 8130 Allison Avenue, La Mesa
- **April 27, 6 p.m. San Marcos City Council Chambers** 1 Civic Center Drive, San Marcos
- **April 28, 6 p.m. Solana Beach Council Chambers** 635 S. Highway 101, Solana Beach
- **May 3, 6 p.m. County Operations Center Annex**, Planning Commission Hearing Room, 5201 Ruffin Road, San Diego

Additionally, from now until May 9, members of the public can submit proposed plans for one or more of the County's five supervisorial districts. In County redistricting, a "plan" means a set of data that assigns Census tracts or blocks to a particular district. 2010 Census data for building these plans is now online, along with detailed instructions, definitions and submission guidelines.

The County's Redistricting Advisory Committee is the citizen's panel charged with gathering public opinion on how to adjust the County's supervisorial district boundaries to balance their populations. Such balancing is required by law every ten years following the Census. Ultimately, the Committee must recommend no more than three proposed redistricting plans to the Board of Supervisors. The Board will adopt a final redistricting plan that takes effect in early September.

Please visit the County's redistricting website at <http://www.sdcounty.gov/redistricting> for more information.

Please feel free to share this information. Also, a one-page flier with general redistricting information is posted online at http://www.sdcounty.ca.gov/redistricting/Resource_documents/RedistrictingInfoFlier.pdf.

Sincerely,

MIKEL HAAS
Deputy Chief Administrative Officer
County of San Diego

Potter, Andrew

From: CSG, Redistricting 2011
Sent: Thursday, April 14, 2011 9:24 AM
To: CSG, Redistricting 2011
Subject: RE: COUNTY REDISTRICTING 2011: IT'S TIME TO BE HEARD

Please note that the correct link for the County's redistricting website is www.sdcounty.ca.gov/redistricting.

From: CSG, Redistricting 2011
Sent: Wednesday, April 13, 2011 4:53 PM
To: CSG, Redistricting 2011
Subject: COUNTY REDISTRICTING 2011: IT'S TIME TO BE HEARD

We want to hear from you! Over the next few weeks, the County of San Diego's Redistricting Advisory Committee will convene in each of the County's five districts in a series of public meetings, giving residents an opportunity to learn about County redistricting and voice your ideas.

At each meeting, County staff will present an overview of the County's redistricting process, including the laws, principles and timelines that guide it. The presentation will review population and demographic data and will explain how the public can submit redistricting plans, give suggestions and stay involved using the County's redistricting website.

The meetings are also opportunities for residents to address the Redistricting Advisory Committee with suggestions, plans and information about your community.

District Meeting Schedule

- **April 20, 6 p.m. Chula Vista Council Chambers**, 276 Fourth Avenue, Chula Vista
- **April 25, 6 p.m. La Mesa City Council Chambers** 8130 Allison Avenue, La Mesa
- **April 27, 6 p.m. San Marcos City Council Chambers** 1 Civic Center Drive, San Marcos
- **April 28, 6 p.m. Solana Beach Council Chambers** 635 S. Highway 101, Solana Beach
- **May 3, 6 p.m. County Operations Center Annex**, Planning Commission Hearing Room, 5201 Ruffin Road, San Diego

Additionally, from now until May 9, members of the public can submit proposed plans for one or more of the County's five supervisorial districts. In County redistricting, a "plan" means a set of data that assigns Census tracts or blocks to a particular district. 2010 Census data for building these plans is now online, along with detailed instructions, definitions and submission guidelines.

The County's Redistricting Advisory Committee is the citizen's panel charged with gathering public opinion on how to adjust the County's supervisorial district boundaries to balance their populations. Such balancing is required by law every ten years following the Census. Ultimately, the Committee must recommend no more than three proposed redistricting plans to the Board of Supervisors. The Board will adopt a final redistricting plan that takes effect in early September.

Please visit the County's redistricting website at <http://www.sdcounty.gov/redistricting> for more information.

Please feel free to share this information. Also, a one-page flier with general redistricting information is posted online at http://www.sdcounty.ca.gov/redistricting/Resource_documents/RedistrictingInfoFlier.pdf.

Sincerely,

MIKEL HAAS
Deputy Chief Administrative Officer
County of San Diego